SUMMARY

THE COMPLETE STORY

Successful replication requires DNA to be copied with high fidelity, failure of which would result in mutations that can be deleterious to the organism. Frequently, mismatches in base pairing occur during DNA replication or homologous recombination and are subsequently repaired by a number of different yet overlapping mechanisms. DNA mismatch repair is responsible for the maintenance of DNA fidelity that lowers the mutation frequency by a factor of 100-1000. Enzyme systems capable of recognition and correction of base pairing errors within the DNA helix have been demonstrated in bacteria, fungi, and mammalian cells. One of the most extensively studied prokaryotic mismatch repair systems is the methyl-directed mismatch repair pathway of *E. coli*. The mismatch repair system of *E. coli* has been reconstituted *in vitro* using purified components and has been shown to require nearly 11 proteins, MutS, MutL, MutH, DNA helicase II, SSB, Rec J, Exonuclease I, Exonuclease VII, Exonuclease X, DNA polymerase III holoenzyme, and DNA ligase. Mutant strains defective in *mutH, mutL, mutS* or *uvrD* are spontaneous mutators and are deficient in mismatch correction.

Haemophilus influenzae is a naturally transformable Gram-negative bacterium, widespread in its distribution among the human population and its pathogenic potential critically depends on the phenomenon of phase variation of number of surface-expressed molecules. The implications of DNA mismatch repair pathway in generating phase

variation, thereby contributing to increased virulence. In *H. influenzae* is under active investigation. In this study, the functional characterization of MutH, MutL and MutS proteins of *H. influenzae* is reported.

Sequencing of the complete genome of *H. influenzae* Rd revealed the existence of the homologs of *E. coli* DNA mismatch repair genes, i.e., *mutH*, *mutL* and *mutS*. None of these gene products of *H. influenzae*, except *mutH*, has been studied functionally in detail so far. It was, therefore envisaged that a biochemical characterization of the mismatch repair proteins would lead to a better understanding of the mechanistic aspects of *H. influenzae* DNA mismatch repair pathway. Hence, this study was initiated with the following objectives: (A) To demonstrate DNA mismatch repair activity in *H. influenzae* cell free extracts, (B) PCR amplification and cloning of *H. influenzae mutH, L, S* genes based on the published genomic DNA sequence; over-expression and purification of the encoded proteins, (C) Complementation assays to evaluate the activities of the *H.influenzae* genes/proteins in corresponding *E. coli* mutants, (D) Characterization of the catalytic and DNA binding properties of *H. influenzae* MutH, (E) Studies on *H. influenzae* MutL and its interaction with MutH, (F) Functional analysis of the ATP-regulated MutS and MutL molecular switches and (G) Studies on the DNA binding properties of *H. influenzae* MutS.

DNA mismatch repair in *H. influenzae* cell free extracts

The deduced amino acid sequence of *H. influenzae* MutH, MutL and MutS showed a high degree of homology with the corresponding *E. coli* proteins, the sequence identity being 58%, 44% and 66%, respectively. Cloning of the *mutH, L, S* genes from *H. influenzae* into pET15b expression vector was performed by PCR using primer sequences derived from the 5'- and 3'- end of the open reading frames obtained from the genomic DNA sequences. The cellular activities of the *H. influenzae* MutL, MutS and MutH proteins were assessed by *in vivo* complementation assays using *mutL*, *mutS* or *mutH* deficient strains of *E. coli*. *H. influenzae* MutH was able to complement the *mutH*-deficient *E. coli* strain back to wild-type level. The extent of complementation obtained for *H. influenzae* MutS was nearly 72% whereas MutL activity in the *E. coli* cell was only 35%. Comparing *E. coli* and *H. influenzae* mismatch repair proteins, MutH and

MutS sequences were found to be highly similar (nearly 77%) whereas MutL showed a similarity of only 58%. This partially explains the differences in the level of complementation by the *H. influenzae* MutH, L, S proteins in an *E. coli* cellular milieu. Repair assays have shown that the cell free extract of *H. influenzae* is proficient in repairing DNA mismatches. *In vitro* complementation assays using a 6440 bp covalently closed circular hemi-methylated GT heteroduplex with the addition of cell free extracts prepared from *mutH* or *L* or *S* deficient strains of *E. coli* and the purified *H. influenzae* MutH, L, S proteins gave similar complementation results.

Subunit composition of *H. influenzae* MutH, L, S proteins

Gel filtration analysis was performed to determine size and subunit structure of purified *H. influenzae* MutL, MutS and MutH proteins in solution. MutL eluted as two peaks, a minor and a major, at positions corresponding to a globular protein of ~80 and ~170 kDa, respectively. This suggests that the protein exists as a dimer in solution under native conditions. The oligomerization state of *H. influenzae* MutS protein in solution under conditions used for heteroduplex DNA binding was analyzed. MutS eluted at a peak corresponding to 372 kDa in addition to giving rise to a few minor peaks corresponding to molecular weights above 400 kDa including one in the void volume. Since the molecular mass of MutS monomer is 99 kDa (including the His_6-tag), these results suggested that MutS protein is present predominantly as a tetramer along with some undefined higher order oligomers, in solution. The purified MutH protein eluted at a peak corresponding to a molecular mass of 29 kDa irrespective of the protein concentration used, indicating that *H. influenzae* MutH is a monomer in solution.

***H. influenzae* MutH: catalytic and DNA binding properties**

H. influenzae MutH endonuclease activity was monitored by the conversion of the covalently closed circular fMR1 DNA (containing a single GATC) to open circular form when incubated with increasing concentrations of the protein (1 nM - 13 μM). 500 ng (6 nM) of the fully unmethylated DNA was completely converted to open circular form at a MutH concentration of approx. 800 nM, when incubated at 37°C for 1 hr. Thus, the weak endonuclease MutH has to be provided at a molar excess of up to 130-fold to completely

convert covalently closed circular DNA to open circular form. On the contrary, only 2-fold molar excess concentration of MutH was capable of converting the covalently closed circular substrate completely to open circular form in the presence of MutL and ATP, compared to 130-fold required in the absence of MutL. The level of activation of MutH by MutL was found to be more than 60-fold compared to the MutH nicking activity in the absence of MutL.

DNA binding properties of *H. influenzae* MutH was studied using electrophoretic gel mobility shift assay using 34 bp DNA duplexes containing a single d(GATC), which was either unmethylated, hemi-methylated or fully methylated. A shifted complex was obtained upon the binding of *H. influenzae* MutH to the hemi-methylated d(GATC) containing duplex. Curiously, similar complexes were also obtained when unmethylated as well as fully methylated DNA duplexes were used in the binding reaction. The binding affinity of MutH was similar for all the d(GATC) duplexes, irrespective of their methylation state. The mobility of the DNA-protein complex appeared to be retarded to varying levels with increasing concentration of MutH. This observation indicated that more than one molecule of MutH binds to each of the duplex DNA molecule. It was surprising to see that MutH binds to the control duplex having a d(GGTC) or d(ACTT) with almost the same affinity as that for hemi-methylated d(GATC). Absence of divalent metal ion in the binding reaction did not alter the affinity of MutH for the duplex DNA. Furthermore, the presence of MutL and ATP in the binding reaction resulted in the formation of a retarded species that showed the same mobility irrespective of the MutH concentration used. This retarded complex appeared to be formed by a single molecule of MutH bound specifically to the d(GATC) duplex.

Mutational analysis of the C-terminal of *H. influenzae* MutH

The mechanism of MutH activation by MutL involves stimulation of both DNA binding and cleavage by MutH. MutL-MutH interaction sites have not been precisely identified so far, experimentally. It has been proposed that the non-conserved C-terminal tail of *E. coli* MutH is important for its interaction with MutL. Clustal W analysis of MutH sequences of homologues identified in the completely sequenced genomes of members of proteobacteria with that of *H. influenzae* helped in the identification of a

conserved cryptic AI/LLAR motif at the C-terminal end of the protein. Site-specific mutagenesis of the coding region of *H. influenzae* MutH protein was carried out resulting in a single amino acid substitution of Ala for Ile 213 and Ala for Leu 214. It was clearly seen that, both the purified I213A and L214A mutants of *H. influenzae* MutH considerably lost their intrinsic DNA nicking activity when compared with that of the wild type protein. The I213A mutant was almost completely inactive whereas the L214A had approx. 30% of residual activity. *In vivo* complementation assays have shown that the I213A mutant was able to complement the *E. coli mutH* defective phenotype only by 80% whereas the L214A mutant gave a higher level of complementation up to 97%.

Similar assays were carried out in the presence of 1 μM MutL and 1 mM ATP to test whether the mutants were amenable to activation by MutL. I213A mutant showed greatly reduced level of activation by MutL compared to L214A mutant. Though the intrinsic endonuclease activity of the L214A mutant was extremely less, the level of activation by MutL was almost comparable with that of the wild type MutH protein. Electrophoretic gel mobility shift assay have shown that the DNA binding affinity of both I213A and L214A mutants of MutH was considerably reduced. The mutagenesis studies of MutH protein indicate that the AI/LLAR motif might be playing an important role at the hinge-region between the two sub-domains of MutH, facilitating active site formation by bringing the two structural domains together.

H. influenzae MutL and MutS: ATP regulated molecular switches

H. influenzae MutL protein exhibited a considerable level of ATPase activity. To determine initial velocities, product formation was measured under conditions such that the final concentration of ADP formed in the reaction was less than 5 %. Using Eadie-Scatchard plot, the kinetic parameters were calculated. The k_{cat} of *H. influenzae* MutL ATPase was found to be 19.2 ± 1.6 min^{-1} and K_m (ATP) was 200 ± 20.5 μM at pH 8.0 and 37°C. Recently, the full length crystal structure of *E. coli* MutL had been reconstructed which clearly indicated protein surface residues that might be involved in DNA binding. Kinetic parameters of MutL ATPase in the presence of single-stranded DNA were determined. The k_{cat} of *H. influenzae* MutL in presence of single-stranded DNA was 51.4 ± 2.1 min^{-1} and K_m for ATP was 425.8 ± 46.5 μM. The k_{cat} in the presence

of single stranded DNA was nearly doubled when compared to assays in the absence of DNA. In addition, K_m for ATP by *H. influenzae* MutL was increased from 200 to 425.8 μM in presence of DNA. This clearly indicated that the protein has reduced affinity for binding ATP when complexed with DNA. Though the ATP binding domain of *H. influenzae* MutL resembles that of DNA gyrase (GHKL ATPase family), its ATPase activity was not inhibited by the presence of novobiocin in the reaction unlike what was observed for the *E. coli* counter part.

The MutS ATPase activity was enhanced approx. 2 fold with linear 35 bp GT heteroduplex and approx. 3 fold with ssDNA. Using Eadie–Scatchard plot, kinetic parameters were calculated. The k_{cat} of *H. influenzae* MutS ATPase was found to be 19.4 ± 0.5 min^{-1} and K_m (ATP) was 359.4 ± 35.7 μM at pH 8.0 and 37 °C. The kinetics of ATP hydrolysis by *H. influenzae* MutS in presence of DNA were also determined. In presence of GT heteroduplex, the MutS ATPase showed a K_m of 375.1 ± 8.2 μM ATP and the maximal hydrolysis rate (k_{cat}) was 37.6 ± 1 min^{-1}. When ssDNA was included in the reaction, the K_m and k_{cat} were 517.6 ± 15.2 μM ATP and 63.2 ± 1.6 min^{-1}, respectively. Interestingly, in the presence of DNA, substrate inhibition at higher concentrations of ATP was observed, which was highly significant when ssDNA was included in the reaction. The functional significance of this observation is not very clear. It could be indicative of the differential regulation of the activities of MutS upon encounter with single-stranded or duplex DNA during the repair process.

More interestingly, there was a 2-fold reduction in the ATPase activity of MutS when covalently closed circular GT heteroduplex was used in the reaction. This general pattern of ATP hydrolysis was observed when MutS and MutL combination was used or MutS, MutL and MutH were used along with the three different DNA substrates. In all the cases, there was a reduction in the overall activity only when covalently closed circular GT heteroduplex was present in the assay reaction. It is possible that, the *H. influenzae* MutS dimers bind to the GT mismatch present in the covalently closed circular GT heteroduplex. After first round of ATP hydrolysis, the MutS complexes might be trapped in the covalently closed circular DNA molecule. These trapped molecular MutS clamps are unable to do the second round of ATP hydrolysis unless and until they

encounter DNA ends or dissociate from the DNA by the intervention of some other protein partners involved in the mismatch repair process. This probably accounts for the reduction in the ATPase activity of MutS when covalently closed circular GT heteroduplex was used in the reaction. Linearisation of the MutS-bound covalently closed circular heteroduplex by digestion with ClaI resulted in the restoration of the ATPase activity to the level that was observed in reactions carried out in the absence of DNA or in presence of homoduplex DNA.

DNA binding properties of *H. influenzae* MutS

From the analysis of the DNA binding data, it can be concluded that the *H. influenzae* MutS exhibited differences in its affinities for the various mismatched bases and the order of affinity observed being GG>GT>ΔT>AA>AC≈CT>TT >GA>CC>GC. The hierarchy of apparent affinities of MutS for at least two mispairs (GT and AC) did correlate well with the efficiencies with which these mispairs were replaced in the *in vitro* assay.

H. influenzae MutS was also found to be capable of binding to single-stranded DNA and the DNA-protein complex could be competed out when excess of unlabelled single-stranded DNA was included in the reaction. This observation of MutS binding to single-stranded DNA was surprising, though its affinity was much less compared to that for GT or GG heteroduplex DNA. It is quite possible that MutS encounters single-stranded DNA regions during the exonucleolytic cleavage step elicited upon recognition of the mismatched base. The *in vivo* significance of this observation at present is unclear.

Conformational changes of *H. influenzae* MutS upon nucleotide binding

In order to understand the influence of DNA and nucleotide on the conformational changes acquired by *H. influenzae* MutS, limited proteolysis pattern by trypsin digestion was monitored. It was quite clear that, in presence of nucleotide and DNA, MutS undergoes drastic conformational change which results in the absence of several proteolytic products in the size range of 25-65 kDa as visualized on the polyacrylamide gels. Circular dichroism spectra of *H. influenzae* MutS in the absence or presence of nucleotide reveals that the protein is structurally disordered. However, the presence of

ATP and heteroduplex DNA conferred MutS with a more ordered structure. This was reflected in the increased ellipticity compared to conditions where MutS and heteroduplex were present without nucleotide or when ADP was included in the sample. The structural perturbation and disorder of MutS were observed when homoduplex DNA was included irrespective of the presence or absence of nucleotide. This clearly indicates that the ordered and compact functional conformation of MutS can be achieved only through the concerted effects of both ATP and heteroduplex DNA.

Taken together, this investigation leads to a better understanding of the structure-function relationships operating among the protein partners involved in the DNA mismatch repair pathway. Unraveling the repair processes that control mutation rates of simple sequences in *H. influenzae* is central to understanding the contribution of phase-variable genes in the host-pathogen relationship. Future studies will be directed towards understanding this mechanism in detail.